Robust Multi-Sensor Classification via Joint Sparse Representation

Nam H. Nguyen
Electrical and Computer Engineering
The Johns Hopkins University
Maryland, U.S.A.
Email: nam@jhu.edu

Nasser M. Nasrabadi
Army Research Lab
Maryland, U.S.A.
Email: nasser.m.nasrabadi@us.army.mil

Trac D. Tran
Electrical and Computer Engineering
The Johns Hopkins University
Maryland, U.S.A.
Email: trac@jhu.edu

Abstract—In this paper, we propose a novel multi-task multi-variate (MTMV) sparse representation method for multi-sensor classification, which takes into account correlations between sensors simultaneously while considering joint sparsity within each sensor's observations. This approach can be seen as the generalized model of multi-task and multivariate Lasso, where all the multi-sensor data are jointly represented by a sparse linear combination of training data. We further modify our MTMV model by including a clutter noise term that is also assume to be sparse in feature domain. An efficient algorithm based on alternative direction method is proposed for both models. Extensive experiments are conducted on real data set and the results are compared with the conventional discriminative classifiers to verify the effectiveness of the proposed methods.

I. INTRODUCTION

Multi-sensor fusion have received considerable amount of attentions over past few years for both military and non-military tasks [1] [2] [3]. A particular interest of multi-sensor fusion is classification, where the ultimate question is how to take advantage of having related information from different sources (sensors) recording the same physical event to achieve improved classification performance. A variety of approaches have been proposed in the literature to answer this question [4] and [5]. These methods mostly fall into two categories: Decision in - decision out (DI-DO) and feature in - feature out (FI-FO) [3]. In [4], the authors investigated the DI-DO method on vehicle classification problem using data collected from acoustic and seismic sensors. They proposed to perform local classification for each sensor signal by conventional methods such as Support Vector Machine (SVM). These local decisions are then incorporated via Maximum A Posterior (MAP) estimator to make the final classification decision, thus named DI-DO method. In [5], FI-FO method is studied for vehicle classification using both visual and acoustic sensors. They proposed a method to extract temporal gait patterns from both sensor signals and utilize them as inputs for SVM classifier. They furthermore compared DI-DO and FI-FO approaches on their dataset and showed the higher discrimination performance of FI-FO over DI-DO.

In signal processing, most natural signals are inherently sparse in certain bases or dictionaries where they can be approximately represented by only a few significant components carrying the most relevant information. In other words, the intrinsic signal information usually lies in a low-dimensional subspace and the semantic information is often encoded in the sparse representation. Especially with the emergence of the Compressed Sensing (CS) framework [6] and [7], sparse representation and related optimization problems involving sparsity as a prior called sparse recovery have increasingly attracted the interest of researchers in various diverse disciplines.

Though the usage of sparsity has been successfully employed in inverse problem, where sparsity acts as a strong prior to abbreviate ill-posedness of the problem. Recent research has pointed out that sparse representation is also useful for discriminative applications [8] [9] [10]. These applications rely on the crucial observation that it is possible to represent the test sample as a linear combination of training samples belonging to the same class as the target and not to the other classes. Thus, if the dictionary is constructed from all the training samples in all the classes, the test samples can be sparsely represented by only a few columns of this dictionary. Therefore, the sparse coefficient vector, which is recovered efficiently via l_1-minimization, can naturally be considered as the discriminative factor. In [9], the authors successfully applied this idea to the face recognition problem. Since then, many more complicated techniques have been exploited and applied to various fields such as hyperspectral target detection [11], acoustic signal classification [12] and visual classification [8] [13] [14]. For instance, the authors of [12] proposed a joint sparse model for acoustic signal classification, which exploits the fact that multiple observations from the same class could be simultaneously represented by few columns of the training dictionary. Thus, coefficient vectors associated with these observations might deliver the same sparse pattern. Similarly, [14] investigated a multi-task model for visual classification, which also assumes tasks belonging to the same class have the same sparse support distributions on coefficient vectors. To improve classification performance, all these models require efficient algorithms that take into account this precious piece of sparsity as a prior information.

In this paper, we consider a multi-sensor classification problem, focusing on discriminating between a human and non-human footstep activity. The experimental setup is as follows: A set of four acoustic, three seismic, one passive infrared (PIR) and one ultrasonic sensors are used to measure

Report Documentation Page

1. REPORT DATE **JUL 2011**	2. REPORT TYPE	3. DATES COVERED **00-00-2011 to 00-00-2011**

4. TITLE AND SUBTITLE **Robust Multi-Sensor Classification via Joint Sparse Representation**	5a. CONTRACT NUMBER
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **U.S. Army Research Laboratory,Adelphi,MD,20783-1197**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES
Presented at the 14th International Conference on Information Fusion held in Chicago, IL on 5-8 July 2011. Sponsored in part by Office of Naval Research and U.S. Army Research Laboratory.

14. ABSTRACT
In this paper, we propose a novel multi-task multivariate (MTMV) sparse representation method for multi-sensor classification, which takes into account correlations between sensors simultaneously while considering joint sparsity within each sensor?s observations. This approach can be seen as the generalized model of multi-task and multivariate Lasso, where all the multi-sensor data are jointly represented by a sparse linear combination of training data. We further modify our MTMV model by including a clutter noise term that is also assume to be sparse in feature domain. An efficient algorithm based on alternative direction method is proposed for both models. Extensive experiments are conducted on real data set and the results are compared with the conventional discriminative classifiers to verify the effectiveness of the proposed methods.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**8**	

same physical event simultaneously on the field. The ultimate goal is to detect whether the event is human or human and animal footsteps. As opposed to the previous approaches on this problem in which only one single sensor is utilized to record data [15] [16] [17], in this paper we propose a novel regularized regression method, namely multi-task multivariate lasso, which effectively incorporates both multi-task and multivariate Lasso ideas [8] and [18]. This technique imposes group sparsities both within each task (sensor) and across the tasks. Furthermore, we extend our model to deal with sparse noise with arbitrarily large magnitude. This high-energy environmental noise frequently appears in sensor data due to the unpredictable or uncontrollable nature of the environment during the data collection process. Though our technique is designed for military purposes, it is not restricted to this specific application. Rather, it can be applied to any set of classification or discrimination problems, where data is usually collected from multiple sources.

The remainder of this paper is organized as follows. Section II briefly introduces various sparsity models with the main focus on our proposed multi-task multivariate (MTMV) and MTMV with sparse noise models in subsections C and D, respectively. We present in Section III a fast and efficient algorithm to solve convex optimizations for these models. Extensive experiments are shown in Section IV and conclusions are drawn in Section V.

II. SPARSITY MODELS

Consider a multi-task (multi-sensor) C-class classification problem. Suppose we have a training set of p samples in which each sample has D different modalities of features or D different tasks. For each task (sensor) $i = 1, ..., D$, we denote $\boldsymbol{X}^i = [\boldsymbol{X}_1^i, \boldsymbol{X}_2^i, ..., \boldsymbol{X}_C^i]$ as a $n \times p$ dictionary, consisting of C sub-dictionaries $\boldsymbol{X}_k^i$'s with respect to C classes. Here, each sub-dictionary $\boldsymbol{X}_j^i = [\boldsymbol{x}_{j,1}^i, \boldsymbol{x}_{j,2}^i, ..., \boldsymbol{x}_{j,p_j}^i] \in \mathbb{R}^{n \times p_j}$ represents a set of training data from the ith task labeled with jth class. Accordingly, $\boldsymbol{x}_{j,k}^i$, which we usually call an atom in the dictionary is the kth training sample for ith task and jth class. Notice that p_j is the number of training sample for class jth and n is the feature dimension of each sample, therefore, the total samples is $p = \sum_{j=1}^C p_j$. Given a test sample $\boldsymbol{Y}$ comprising of D tasks $\{\boldsymbol{Y}^1, \boldsymbol{Y}^2, ..., \boldsymbol{Y}^D\}$ where each sample task $\boldsymbol{Y}^i$ consists of d_i observations $\boldsymbol{Y}^i = [\boldsymbol{y}_1^i, \boldsymbol{y}_2^i, ..., \boldsymbol{y}_{d_i}^i] \in \mathbb{R}^{n \times d_i}$, we want to decide which class the sample $\boldsymbol{Y}$ belongs to.

A. Sparse representation for classification (SRC)

We first review the single task (single sensor) sparse representation for classification method. Accordingly, we remove the subscript i representing tasks for simplicity. In this problem, a particular and effective model is to assume that the training samples belonging to the same class approximately lie on a low-dimensional subspace. Given a set of C distinct classes $\boldsymbol{X} = [\boldsymbol{X}_1, \boldsymbol{X}_2, ..., \boldsymbol{X}_C]$ where jth class has p_j training samples $\{\boldsymbol{x}_{j,k}\}_{k=1,...,p_j}$, a new test sample $\boldsymbol{y}$ belonging to the jth class will approximately lie in the linear space spanned by

the training samples associated with jth class:

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{w} + \boldsymbol{n}, \tag{1}$$

where $\boldsymbol{w}$ is the coefficient vector whose entries have value 0's except those associated with the jth class: $\boldsymbol{w} = [\boldsymbol{0}^T, ..., \boldsymbol{0}^T, \boldsymbol{w}_j^T, \boldsymbol{0}^T, ..., \boldsymbol{0}^T]^T$, and $\boldsymbol{n}$ is a small noise due to the imperfectness of the test sample.

In order to obtain the sparse vector $\boldsymbol{w}$, it is natural to consider the following optimization

$$\widehat{w} = \underset{\boldsymbol{w}}{\arg\min} \|\boldsymbol{w}\|_0 \qquad \text{subject to} \qquad \|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{w}\|_2 \le \epsilon, \tag{2}$$

where $\|\boldsymbol{w}\|_0$ is l_0-norm defined as the number of non-zero entries of $\boldsymbol{w}$ and ϵ is the noise energy. This l_0-norm minimization can be interpreted as finding the sparsest solution obeying the quadratic constraints. However, (2) is well-known as an NP-hard problem due to the non-convexity and non-differentiability of the l_0-norm. Many alternative approaches have been proposed to approximately solve (2) such as greedy pursuits (Orthogonal Matching Pursuit [19], Subspace Pursuit [20]) and Iterative Hard Thresholding (IHT) [21]. Alternatively, l_0-minimization can be efficiently solved by recasting it as l_1-based convex programming problem [6] [7]. In this paper, we utilize the l_1-minimization approach which is described as follows

$$\widehat{\boldsymbol{w}} = \underset{\boldsymbol{w}}{\arg\min} \frac{1}{2} \|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{w}\|_2^2 + \lambda \|\boldsymbol{w}\|_1, \tag{3}$$

where λ is a positive regularization and l_1-norm $\|\boldsymbol{w}\|_1 = \sum_{i=1}^p |w_i|$. This optimization is also known as Lasso [22], which can be solved efficiently in polynomial time by standard convex optimization techniques.

Once the coefficient vector $\widehat{\boldsymbol{w}}$ is estimated, the class label of $\boldsymbol{y}$ is determined by the minimal residue between $\boldsymbol{y}$ and its approximation from each class sub-dictionary

$$\widehat{j} = \underset{j}{\arg\min} \|\boldsymbol{y} - \boldsymbol{X}\delta_j(\boldsymbol{w})\|_2, \tag{4}$$

where $\delta_j(\cdot)$ is a vector indicator function, defined by keeping the coefficients corresponding to the jth class and setting all others to be zero, i.e. $\delta_k(w^i) = [0^T, ..., 0^T, w_j^T, 0^T, ..., 0^T]^T$.

B. Multivariate sparse representation for classification (MV-SRC)

Let us first consider a single task sparse representation where the test sample is generated by a single sensor. However, the test sample $\boldsymbol{Y}$ may consists of multiple observations of the same physical event obtained by the same sensor: $\boldsymbol{Y} = [\boldsymbol{y}_1, \boldsymbol{y}_2, ..., \boldsymbol{y}_d] \in \mathbb{R}^{n \times d}$. In our problem, each observation is one segment of the test signal where each segment is obtained by partitioning the test signal into d (overlapping) segments. Again, suppose the test signal belongs to jth class, it can be often assumed that each observation $\boldsymbol{y}_i$ is a linear combination of training samples in the sub-dictionary $\boldsymbol{X}_j$ which consists of training segments. That is, for all $i = 1, ..., d$, $\boldsymbol{y}_i = \boldsymbol{X}\boldsymbol{w}_i + \boldsymbol{n}_i$ where $\boldsymbol{X} = [\boldsymbol{X}_1, \boldsymbol{X}_2, ..., \boldsymbol{X}_C]$ is a concatenation of C sub-dictionaries and $\boldsymbol{w}_i$'s are sparse

vectors whose nonzero entries are associated with the jth class: $\boldsymbol{w}_i = [\boldsymbol{0}^T, ..., \boldsymbol{0}^T, \boldsymbol{w}_{i,j}^T, \boldsymbol{0}^T, ..., \boldsymbol{0}^T]^T$ and $\boldsymbol{n}_i$'s are small noises. If we denote $\boldsymbol{W} = [\boldsymbol{w}_1, \boldsymbol{w}_2, ..., \boldsymbol{w}_d] \in \mathbb{R}^{n \times d}$, then $\boldsymbol{W}$ is a row-sparse matrix with only p_j nonzero rows located at the jth class.

To recover the row-sparse matrix $\boldsymbol{W}$, the following joint sparse optimization is proposed with $q \geq 1$. This optimization has been well known as group Lasso [18] [23].

$$\widehat{\boldsymbol{W}} = \operatorname*{argmin}_{\boldsymbol{W}} \frac{1}{2} \|\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{W}\|_F^2 + \lambda \|\boldsymbol{W}\|_{1,q}, \qquad (5)$$

where λ is a positive regularization parameter, and $\|\boldsymbol{W}\|_{1,q}$ is a norm defined as $\|\boldsymbol{W}\|_{1,q} = \sum_{k=1}^{p} \|\boldsymbol{w}^k\|_q$ where $\boldsymbol{w}^k$'s are row vectors of $\boldsymbol{W}$. This norm can be phrased as performing l_q-norm cross the columns (observations) and then l_1-norm along rows. It is clear that this l_1/l_q regularization norm encourages shared patterns across related observations, and thus solution of the optimization (5) has common support at column level.

The class label is determined by the following rule

$$\widehat{j} = \operatorname*{argmin}_{j} \|\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\delta}_j(\boldsymbol{W})\|_F, \qquad (6)$$

where $\boldsymbol{\delta}_j(\cdot)$ is the matrix indication function, defined by keeping rows corresponding to the jth class and setting all others to be zeros.

C. Multi-task multivariate sparse representation for classification (MTMV-SRC)

In the previous section, we have employed a single source sparse representation for classification. In the scenario where an event is captured by multiple heterogeneous sources (sensors), thus multiple observations are available in the test sample. By exploiting correlation between different sources, we can potentially improve classification accuracy. To handle multiple sources, a naive approach is to utilize voting scheme (or DI-DO method), where for each sensor the aforementioned two-step classification algorithm described in Section II-B is performed and a class label is assigned. The final decision is made by selecting the label that occurs the most. It is clear that this approach does not exploit the relationship between different sources except at the post-processing step where decision is made via fusion.

In this section, an alternative approach is proposed in which we exploit a joint sparsity of coefficient vectors from different sources in order to make a joint classification decision. To illustrate this model, let us first consider a two-task classification with the test sample $\boldsymbol{Y}$ consisting of two tasks $\boldsymbol{Y}^1$ and $\boldsymbol{Y}^2$ collected from 2 different sensors. Suppose that $\boldsymbol{Y}^1$ belongs to the jth class, it can be reconstructed by a linear combination of the atoms in the sub-dictionary $\boldsymbol{X}_j^1$. That is, $\boldsymbol{Y}^1 = \boldsymbol{X}^1\boldsymbol{W}^1 + \boldsymbol{N}^1$ where $\boldsymbol{W}^1$ is a sparse matrix with only p_j nonzero rows associated with jth class and $\boldsymbol{N}^1$ is a small noise matrix.

Since $\boldsymbol{Y}^2$ represents the same event, it belongs to the same class, and thus can be approximated by training samples in $\boldsymbol{X}_j^2$ with a different set of coefficients $\boldsymbol{W}_j^2$, $\boldsymbol{Y}^2 = \boldsymbol{X}^2\boldsymbol{W}^2 + \boldsymbol{N}^2$

where $\boldsymbol{W}^2$ has the same sparsity pattern as $\boldsymbol{W}^1$.

If we denote $\boldsymbol{W} = [\boldsymbol{W}^1, \boldsymbol{W}^2]$, then $\boldsymbol{W}$ is a sparse matrix with only p_j nonzero rows. Therefore, in order to seek for this row-sparse matrix, we should incorporate this common sparse pattern prior into the optimization algorithm. In the more general case where we have D sources (sensors), if we denote $\{\boldsymbol{Y}^i\}_{i=1}^{D}$ as a set of D observations each consisting of d segments collected from D sensors and let $\boldsymbol{W} \in \mathbb{R}^{n \times pD}$ be an unknown matrix formed by concatenating coefficient matrices $\boldsymbol{W} = [\boldsymbol{W}^1, \boldsymbol{W}^2, ..., \boldsymbol{W}^D]$. This matrix $\boldsymbol{W}$ can be recovered by solving the following l_1/l_q-regularized least square problem

$$\widehat{\boldsymbol{W}} = \operatorname*{argmin}_{\boldsymbol{W}} \frac{1}{2} \sum_{i=1}^{D} \left\|\boldsymbol{Y}^i - \boldsymbol{X}^i\boldsymbol{W}^i\right\|_F^2 + \lambda \|\boldsymbol{W}\|_{1,q}, \quad (7)$$

where λ is a positive parameter and q is set to be greater than 1 to make the optimization convex. This optimization (7) is called multi-task multivariate Lasso.

Once $\widehat{\boldsymbol{W}}$ is obtained, the class label is decided by minimal residual rule

$$\widehat{j} = \operatorname*{argmin}_{j} \sum_{i=1}^{D} \left\|\boldsymbol{Y}^i - \boldsymbol{X}^i\boldsymbol{\delta}_j^i(\boldsymbol{W}^i)\right\|_F^2, \qquad (8)$$

where δ_j^i is the matrix indication function associated with ith sensor, defined similarly as the aforementioned section.

D. Multi-task multivariate sparse representation with sparse noise (MTMV-SRC+N)

During the process of collecting data in the field, there are many environmental clutter noises such as impulsive noise or wind noise affecting the true characteristics of the signal. Unfortunately, due to the imperfectness of the environment, these noise sources are uncontrollable and can have arbitrarily large magnitude, which sometime dominate the collected signal. It is obvious that by removing these clutter noises it is possible to improve overall classification performance. Fortunately, these type of noises usually occur in certain frequency bands. We expect that this will only affect some coefficients in our cepstral feature domain, which is the feature space used in our experiments. In this case, the linear model of the observation $\boldsymbol{Y}^i$, $i = 1, ..., D$ with respect to the training data $\boldsymbol{X}^i$ should be modified as

$$\boldsymbol{Y}^i = \boldsymbol{X}^i\boldsymbol{W}^i + \boldsymbol{E}^i + \boldsymbol{N}^i, \qquad i = 1, ..., D, \qquad (9)$$

where $\boldsymbol{N}^i$ is a small dense additive noise and $\boldsymbol{E}^i \in \mathbb{R}^{n \times d_i}$ is a matrix of clutter noise with arbitrarily large magnitude. The nonzero entries of $\boldsymbol{E}^i \in \mathbb{R}^{n \times d_i}$ represents which cepstral features of $\boldsymbol{Y}^i$ are corrupted. Note that it might be possible that all these cepstral features are corrupted. The location of corruption can differ for different tasks since sensors with different characteristics will have different types of errors. In a more general scenario, one can assume that each error $\boldsymbol{E}^i$ is sparsely represented with respect to some basis $\boldsymbol{T}^i \in \mathbb{R}^{n \times m^i}$. That is, $\boldsymbol{E}^i = \boldsymbol{T}^i\boldsymbol{Z}^i$ for some sparse matrices $\boldsymbol{Z}^i \in \mathbb{R}^{m_i \times d_i}$.

The idea of exploiting sparse prior of the error has been developed by Wright et. al. [9] in the context of face recognition,

and Candès et. al. [24] in robust principle component analysis. In this section, we propose a new sparse representation method that simultaneously performs classification and removes clutter noise. By taking the advantage of knowing that errors $\boldsymbol{E}^i$ are sparse, we propose to solve the following optimization to retrieve coefficients $\boldsymbol{W}^i$ as well as errors $\boldsymbol{E}^i$

$$(\widehat{\boldsymbol{W}}, \widehat{\boldsymbol{E}}) = \underset{\boldsymbol{W},\boldsymbol{E}}{\operatorname{argmin}} \frac{1}{2} \sum_{i=1}^{D} \left\| \boldsymbol{Y}^i - \boldsymbol{X}^i \boldsymbol{W}^i - \boldsymbol{E}^i \right\|_F^2$$
$$+ \lambda_w \left\| \boldsymbol{W} \right\|_{1,q} + \lambda_e \left\| \boldsymbol{E} \right\|_1, \quad (10)$$

where λ_w and λ_e are positive parameters, and $\boldsymbol{E}$ is a matrix formed by error matrices $\boldsymbol{E}^i$'s: $\boldsymbol{E} = [\boldsymbol{E}^1, \boldsymbol{E}^2, ..., \boldsymbol{E}^D]$. The l_1-norm of matrix $\boldsymbol{E}$ is defined as the sum of absolute value of the entries: $\|\boldsymbol{E}\|_1 = \sum_{ij} |e_{ij}|$. It is clear from this minimization that we impose both common sparsity on $\boldsymbol{W}$ and entry-wise sparsity on the error $\boldsymbol{E}$.

Once the sparse solution $\boldsymbol{W}$ and error $\boldsymbol{E}$ are computed, the clean cepstral features $\boldsymbol{Y}_c^i$ is recovered by setting $\boldsymbol{Y}_c^i = \boldsymbol{Y}^i - \boldsymbol{E}^i$. To identify the class, we slightly modify the label inference in (8) that accounts for the error $\boldsymbol{E}^i$

$$\widehat{j} = \underset{j}{\operatorname{argmin}} \sum_{i=1}^{D} \left\| \boldsymbol{Y}^i - \boldsymbol{X}^i \boldsymbol{\delta}_j^i(\boldsymbol{W}^i) - \boldsymbol{E}^i \right\|_F^2. \quad (11)$$

III. ALGORITHM

In this section, we propose a fast algorithm to solve (10). The optimization (7) can be solved similarly by setting the parameter λ_e in (10) with respect to error regularization to zero. Our algorithm is relied on the classical alternating direction method of multipliers (ADMM). This method has been recently applied successfully into l_1-norm minimization [25] [26].

Denote the loss function $\mathcal{L}(\boldsymbol{W}, \boldsymbol{E}) = \frac{1}{2} \sum_{i=1}^{D} \left\| \boldsymbol{Y}^i - \boldsymbol{X}^i \boldsymbol{W}^i - \boldsymbol{E}^i \right\|_2^2$, our ultimate goal is to solve the following optimization

$$\min_{\boldsymbol{W},\boldsymbol{E}} \mathcal{L}(\boldsymbol{W}, \boldsymbol{E}) + \lambda_e \left\| \boldsymbol{E} \right\|_1 + \lambda_w \left\| \boldsymbol{W} \right\|_{1,q}. \quad (12)$$

One of the key ideas of the algorithm is to decouple $\mathcal{L}(\boldsymbol{W}, \boldsymbol{E})$, $\|\boldsymbol{E}\|_1$ and $\|\boldsymbol{W}\|_{1,q}$. This can be performed by introducing auxiliary variables to reformulate the problem into a constrained optimization

$$\min_{\boldsymbol{W},\boldsymbol{E},\boldsymbol{V},\boldsymbol{U}} \mathcal{L}(\boldsymbol{W}, \boldsymbol{E}) + \lambda_e \left\| \boldsymbol{U} \right\|_1 + \lambda_w \left\| \boldsymbol{V} \right\|_{1,q}$$
$$\text{subject to} \quad \boldsymbol{W} = \boldsymbol{V}, \boldsymbol{E} = \boldsymbol{U}. \quad (13)$$

The reason behind variable splitting method is that it might be easier to solve the constrained problem (13) than its unconstrained counterpart (12). Since (13) is an equality constrained problem, the Augmented Lagrangian method (ALM) can be used to solve by minimizing the augmented Lagrangian function $f_{\beta_E, \beta_W}(\boldsymbol{W}, \boldsymbol{E}, \boldsymbol{V}, \boldsymbol{U}; \boldsymbol{B}_E, \boldsymbol{B}_W)$ defined as

$$\mathcal{L}(\boldsymbol{W}, \boldsymbol{E}) + \lambda_e \left\| \boldsymbol{U} \right\|_1 + \langle \boldsymbol{B}_E, \boldsymbol{E} - \boldsymbol{U} \rangle + \frac{\beta_E}{2} \left\| \boldsymbol{E} - \boldsymbol{U} \right\|_F^2$$
$$+ \lambda_w \left\| \boldsymbol{V} \right\|_{1,q} + \langle \boldsymbol{B}_W, \boldsymbol{W} - \boldsymbol{V} \rangle + \frac{\beta_W}{2} \left\| \boldsymbol{W} - \boldsymbol{V} \right\|_F^2, \quad (14)$$

where $\boldsymbol{B}_E$ and $\boldsymbol{B}_W$ are the multipliers of the two linear constraints, and β_E, β_W are the positive penalty parameters. The ALM consists in solving $f_{\beta_E, \beta_W}(\boldsymbol{W}, \boldsymbol{E}, \boldsymbol{V}, \boldsymbol{U}; \boldsymbol{B}_E, \boldsymbol{B}_W)$ with respect to $\boldsymbol{W}$, $\boldsymbol{E}$, $\boldsymbol{U}$ and $\boldsymbol{V}$ jointly, keeping $\boldsymbol{B}_E$ and $\boldsymbol{B}_W$ fixed, and then updating $\boldsymbol{B}_E$ and $\boldsymbol{B}_W$. However, this minimization is often not easy. Fortunately, by considering the separable structure of the objective function f_{β_E, β_W}, we can further simplify the problem by minimizing f_{β_E, β_W} with respect to variables $\boldsymbol{W}$, $\boldsymbol{E}$, $\boldsymbol{U}$ and $\boldsymbol{V}$ separately. The method is called alternating direction method of multipliers (ADMM) and given below

---

1: Choose $\boldsymbol{W}^0$, $\boldsymbol{U}^0$, $\boldsymbol{V}^0$, $\boldsymbol{B}_E^0$ $\boldsymbol{B}_V^0$ and β_E, β_W
2: **While** not converged **do**
3:     $\boldsymbol{W}_{t+1} = \operatorname{argmin}_{\boldsymbol{W}} f(\boldsymbol{W}, \boldsymbol{E}_t, \boldsymbol{U}_t, \boldsymbol{V}_t; \boldsymbol{B}_{U,t}, \boldsymbol{B}_{V,t})$
4:     $\boldsymbol{E}_{t+1} = \operatorname{argmin}_{\boldsymbol{E}} f(\boldsymbol{W}_{t+1}, \boldsymbol{E}, \boldsymbol{U}_t, \boldsymbol{V}_t; \boldsymbol{B}_{U,t}, \boldsymbol{B}_{V,t})$
5:     $\boldsymbol{U}_{t+1} = \operatorname{argmin}_{\boldsymbol{U}} f(\boldsymbol{W}_{t+1}, \boldsymbol{E}_{t+1}, \boldsymbol{U}, \boldsymbol{V}_t; \boldsymbol{B}_{U,t}, \boldsymbol{B}_{V,t})$
6:     $\boldsymbol{V}_{t+1} = \operatorname{argmin}_{\boldsymbol{V}} f(\boldsymbol{W}_{t+1}, \boldsymbol{E}_{t+1}, \boldsymbol{U}_{t+1}, \boldsymbol{V}; \boldsymbol{B}_{U,t}, \boldsymbol{B}_{V,t})$
7:     $\boldsymbol{B}_{U,t+1} := \boldsymbol{B}_{U,t} + \beta_E(\boldsymbol{W}_{t+1} - \boldsymbol{U}_{t+1})$
8:     $\boldsymbol{B}_{V,t+1} := \boldsymbol{B}_{V,t} + \beta_W(\boldsymbol{W}_{t+1} - \boldsymbol{V}_{t+1})$.

Algorithm 1: HMT representation via ADMM

The first optimization subproblem with respect to $\boldsymbol{W}$ (line 3 of the Algorithm 1) has quadratic structure, thus, easy to solve via setting the first-order derivative to zero. Furthermore, since the loss function $\mathcal{L}(\boldsymbol{W}, \boldsymbol{E})$ is a sum of convex functions associated with sub-matrices $\boldsymbol{W}^i$, one can simultaneously seek for $\boldsymbol{W}_{t+1}^i$ which has explicit solution as follows

$$\boldsymbol{W}_{t+1}^i = (\boldsymbol{X}^{i^T} \boldsymbol{X}^i + \beta_W \boldsymbol{I})^{-1} [\boldsymbol{X}^{i^T}(\boldsymbol{Y}^i - \boldsymbol{E}_t^i) + \beta_W \boldsymbol{V}_t^i - \boldsymbol{B}_{V,t}^i], \quad (15)$$

where $\boldsymbol{I}$ is an identity matrix of size $p \times p$, and $\boldsymbol{U}_t^i, \boldsymbol{B}_{U,t}^i, \boldsymbol{V}_t^i, \boldsymbol{B}_{V,t}^i$ are sub-matrices of $\boldsymbol{U}_t, \boldsymbol{B}_{U,t}, \boldsymbol{V}_t$ and $\boldsymbol{B}_{V,t}$, respectively.

The second optimization subproblem with respect to $\boldsymbol{E}$ (line 4 of the Algorithm 1) has similar structure and can be solved by

$$\boldsymbol{E}_{t+1}^i = (1 + \beta_E)^{-1}(\boldsymbol{Y}^i - \boldsymbol{X}^i \boldsymbol{W}_{t+1}^i + \beta_E \boldsymbol{U}_t^i - \boldsymbol{B}_{E,t}^i). \quad (16)$$

The third optimization subproblem with respect to $\boldsymbol{U}$ (line 5 of the Algorithm 1) is the standard l_1-minimization. In fact, by a simple calculus, this optimization is equivalent to

$$\min_{\boldsymbol{U}} \frac{1}{2} \left\| \boldsymbol{E}_{t+1} + \frac{1}{\beta_E} \boldsymbol{B}_{E,t} - \boldsymbol{U} \right\|_F^2 + \frac{\lambda_e}{\beta_E} \left\| \boldsymbol{U} \right\|_1, \quad (17)$$

which is well-known shrinkage problem. The explicit solution is derived by

$$\boldsymbol{U}_{t+1} = \operatorname{Shrink}(\boldsymbol{E}_{t+1} + \frac{1}{\beta_E} \boldsymbol{B}_{E,t}, \lambda_e/\beta_E), \quad (18)$$

where the operator $\operatorname{Shrink}(\cdot)$ performs entry-wise and is defined by $\operatorname{Shrink}(z, \epsilon) = \operatorname{sgn}(z)(|z| - \epsilon)$ for $|z| \geq \epsilon$ and zero otherwise.

The last optimization subproblem with respect to $\boldsymbol{V}$ (line 6

of the Algorithm 1) can be recast as

$$\min_{\boldsymbol{V}} \frac{1}{2} \left\| \boldsymbol{W}_{t+1} + \frac{1}{\beta_W} \boldsymbol{B}_{W,t} - \boldsymbol{V} \right\|_F^2 + \frac{\lambda_w}{\beta_W} \|\boldsymbol{V}\|_{1,q}. \quad (19)$$

It is clear that the objective function in (19) has separable structure. Therefore, this optimization can be tackled by minimizing with respect to each row of $\boldsymbol{V}$ separately. Specifically, if we denote $\boldsymbol{w}_{i,t+1}$, $\boldsymbol{b}_{W,i,t}$ and $\boldsymbol{v}_{i,t+1}$ as rows of matrices $\boldsymbol{W}_{t+1}, \boldsymbol{B}_{W,t}$ and $\boldsymbol{V}_{t+1}$, respectively, then for each $i = 1, ..., p$ we solve a sub-problem,

$$\boldsymbol{v}_{i,t+1} = \underset{\boldsymbol{v}}{\operatorname{argmin}} \frac{1}{2} \|\boldsymbol{z} - \boldsymbol{v}\|_{l_2}^2 + \gamma \|\boldsymbol{v}\|_q, \quad (20)$$

where $\boldsymbol{z} := \boldsymbol{w}_{i,t+1} - \boldsymbol{b}_{W,i,t}/\beta_W$ and $\gamma := \lambda_w/\beta_W$. Though (20) is sufficiently simple to derive an explicit solution for any value of q. In this paper we only consider the simplest situation $q = 2$. It is now easy to check that solution of (20) has explicit form

$$\boldsymbol{v}_{i,t+1} = \left(1 - \frac{\gamma}{\|\boldsymbol{z}\|_2} \right)_+ \boldsymbol{z}$$

where $(\boldsymbol{u})_+$ is a vector with entries $\max(u_i, 0)$.

IV. SIMULATION RESULTS AND ANALYSIS

In this section, we perform extensive experiments on real multi-sensor data sets and compare the results with several conventional classification methods such as logistic regression and SVM to verify the effectiveness of our proposed approach.

A. Experiment setup

In this part, we explain briefly the experiment setup.

1. Data collection. Footstep data collection was conducted by two sets of nine sensors consisting of four acoustic, three seismic, one passive infrared (PIR) and one ultrasonic sensors in two days (see Fig. 1 for 3 different types of sensors). Tests with human footstep include one person walking, one person jogging, two people walking, two people running, and a group of multiple people walking, running. Tests with human-animal footstep include one person leading horse or dog, two people leading a horse and a mule, three people leading a horse, a mule and a donkey and a group of multiple people with several dogs. In each test, people and animal could be carrying varying amount of loads such as backpack, metal pipe. People in the test comprise of both males and females.

During each run, test subjects are asked to follow a path where two sets of sensors are positioned and return to the start point. The two sensor sets are placed separately in which each set consists of all nine sensors. A total of 69 round-trip runs were conducted in two days, including 34 runs for human footstep and 35 runs for human-animal footsteps. The collected data, named DEC09 and DEC10 corresponding to two days December 09 and 10, is presented in Table IV-A.

2. Segmentation. To accurately perform classification, it is necessary to extract the actual events from the run series. Although the raw signal of each run might be several minutes in length, the event is much shorter as it occurs in a short

Data	Human	Human lead animal
DEC09	16	15
DEC10	18	20

Fig. 1. Four acoustic sensors (top left), seismic sensor (top right) and passive infrared (PIR) sensor (bottom).

period of time when the test subject is close to the sensors. In addition, event can be at arbitrary locations. To extract useful features, we need to detect time locations where the physical event occurs. To do this, we identify the location with strongest signal response by spectral maximum detection method [27]. From this location, 10 segments with 75% overlap on both sides of the signals are taken, each has 30000 samples corresponding to 3 seconds signal. This process is performed for all sensor data. Overall, for each run, we have 9 signals captured by 9 sensors, each signal is divided into 10 overlapping segments, thus $D = 9$ and each $d_i = 10$, $i = 1, ..., D$.

Fig. 2 visually demonstrates signals captured by four distinct sensors where the event is one person walking. As one can see, different sensors characterize different signal behaviors. The seismic signal shows more clearly the cadences of the test person while we are not able to observe this event by other sensing signals. To have a closer look at the sensing signals, we further show in Fig. 3 one segment extracted from each of the 9 distinct sensing signals. In this figure, the forth acoustic signal is corrupted due to the sensor failure during collection process.

3. Feature extraction. After segmentation, we extract Cepstral features [28] in each segment and keep the first 500 coefficients for classification. Cepstral features have been proved to be effective in speech recognition and acoustic signal classification. The feature dimension which is represented by the number of extracted cepstral features is $n = 500$.

B. Two class problem

First, we demonstrate the effectiveness of exploiting correlations between multiple sensors over the use of a single sensor on two-class classification problem, in particular classifying human and human-animal footsteps. In this experiment, we use the DEC10 data for training and DEC09 data for testing, which leads to 36 training and 31 testing samples. For each ith sensor, the corresponding training dictionary $\boldsymbol{X}^i$ is constructed from all the cepstral feature segments extracted from the 36 training signals. In our experiments, 10 segments are taken from each individual sensor signal. Therefore, each training dictionaries $\boldsymbol{X}^i$, $i = 1, ..., 9$ is of size 500×360 and the
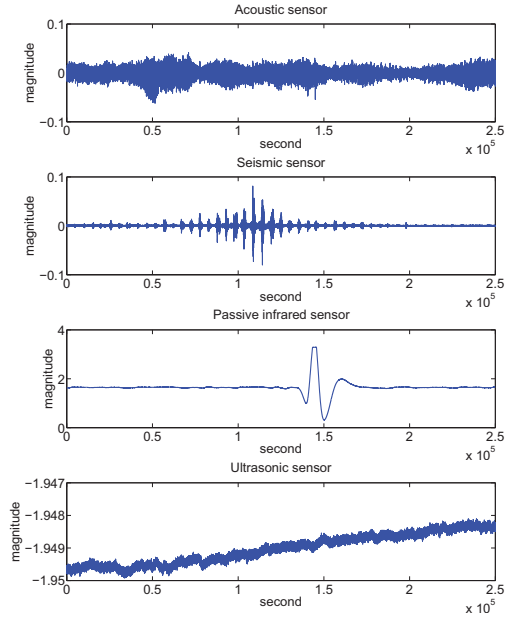
449

Fig. 2. Signals captured by four sensors including acoustic, seismic, PIR and ultrasonic sensors.

associated observations $\boldsymbol{Y}^i$ is of size 500×10 where 500 is the feature dimension.

Classification performance is summarized in Table II, where the first column refers to the methods used in our experiments which include multivariate sparse representation for classification (MV-SRC) for all 9 sensors separately and multi-task MV-SRC (MTMV-SRC) for different combinations of sensors. We note here that the first four sensors are acoustic, the next three (sensors 5 to 7) are seismic sensors and the last two are PIR and ultrasonic sensors, respectively. The second and third columns describe classification accuracy of human and human-animal footsteps, and the last column is the overall accuracy. As can be seen for the table II, MTMV-SRC using the first two acoustic sensors simultaneously outperforms MV-SRC when using each sensor separately. Similar behavior can be observed with three seismic sensors. When all 9 sensors are employed, MTMV-SRC yields the best performance.

It was noticed that during experimentation that half of the testing data collected from two acoustic sensors 3 and 4 in DEC09 is completely noisy due to the malfunction of these two sensors in December 09 (see Fig. 3 for demonstration of a noisy segment extracted from 4th acoustic sensor). This explains why classification performance of the two sensors are quite low compared to sensors 1 and 2.

The next experiment compares our proposed approach with current state-of-the-art classification methods such as (group) sparse logistic regression (SLR) [29], kernel (group) SLR [30], linear support vector machine (SVM) and kernel SVM [31]. In this experiment, 7 sensors, namely sensors $1, 2$ 5-9 are utilized. To exploit correlation across tasks (sensors) in logistic regression method, we utilize the heterogeneous model proposed in [30]. The main idea is to associate each training
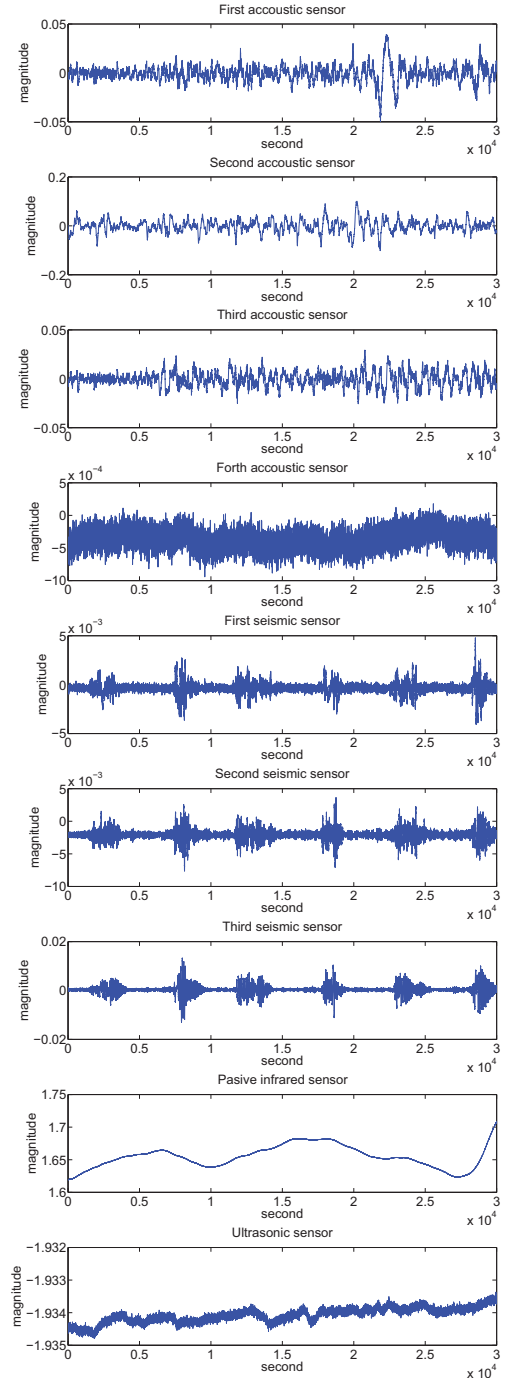

Fig. 3. Signal segments of length 30000 captured by sets of 9 sensors including acoustic, seismic, PIR and ultrasonic sensors.

dictionary $\boldsymbol{X}^i \in \mathbb{R}^{n \times p}$ with a coefficient vector $\boldsymbol{w}^i \in R^n$ and the logistic loss is taken over the sum of all the tasks. Lasso or group Lasso regularization can be incorporated into the optimization to retrieve coefficient vectors $\boldsymbol{w}^i$, $i = 1, ..., 7$. Each segment of the test sample is then assigned to a class and the final decision is made by selecting the label that

TABLE II

CLASSIFICATION ACCURACY (%) FOR TWO CLASS PROBLEM WITH
TRAINING SAMPLES TAKEN FROM DEC10, COMPARING MV-SRC TO
MTMV-SRC WITH DIFFERENT COMBINATIONS OF SENSORS.

Methods	H	HA	OA
MV-SRC sensor 1	75.00	46.67	60.84
MV-SRC sensor 2	75.00	53.33	64.17
MV-SRC sensor 3	37.50	73.33	55.42
MV-SRC sensor 4	75.00	33.33	54.17
MV-SRC sensor 5	50.00	66.67	58.33
MV-SRC sensor 6	56.25	66.67	61.46
MV-SRC sensor 7	56.25	66.67	61.46
MV-SRC sensor 8	31.25	**73.33**	52.29
MV-SRC sensor 9	68.75	53.33	61.04
MTMV-SRC sensors 1-2	**81.25**	60.00	70.63
MTMV-SRC sensors 5-7	75.00	66.67	70.84
MTMV-SRC sensors 1,2,5-9	81.25	66.67	**73.96**

TABLE III

CLASSIFICATION ACCURACY (%) FOR TWO CLASS PROBLEM WITH
TRAINING SAMPLES TAKEN FROM DEC10, COMPARING MTMV-SRC
WITH CONVENTIONAL CLASSIFIERS. 7 SENSORS 1, 2 AND 5-9 ARE USED
IN THIS EXPERIMENT.

Methods	H	HA	OA
MTMV-SRC	81.25	66.67	**73.96**
SLR	81.25	53.33	67.29
Group SLR	75.00	**66.67**	70.84
Kernel SLR	**87.50**	46.67	67.09
Kernel group SLR	**87.50**	53.33	70.42
SVM	81.25	53.33	67.29
Kernel SVM	81.25	60.00	70.63

TABLE IV

CLASSIFICATION ACCURACY (%) FOR TWO CLASS PROBLEM WITH
TRAINING SAMPLES TAKEN FROM DEC09, COMPARING MV-SRC TO
MTMV-SRC WITH DIFFERENT COMBINATIONS OF SENSORS.

Methods	H	HA	OA
MV-SRC sensor 1	66.67	40.00	53.34
MV-SRC sensor 2	**88.89**	35.00	61.95
MV-SRC sensor 5	50.00	85.00	67.50
MV-SRC sensor 6	66.67	65.00	65.84
MV-SRC sensor 7	27.78	70.00	48.89
MV-SRC sensor 8	**88.89**	25.00	56.95
MV-SRC sensor 9	66.67	20.00	43.34
MTMV-SRC sensors 1-2	94.44	35.00	64.72
MTMV-SRC sensors 5-7	38.89	**95.00**	66.95
MTMV-SRC sensors 1,2,5-9	77.78	65.00	**71.39**

TABLE V

CLASSIFICATION ACCURACY (%) FOR TWO CLASS PROBLEM WITH
TRAINING SAMPLES TAKEN FROM DEC09, COMPARING MTMV-SRC
WITH CONVENTIONAL CLASSIFIERS. 7 SENSORS 1, 2 AND 5-9 ARE USED
IN THIS EXPERIMENT.

Methods	H	HA	OA
MTMV-SRC	77.78	65.00	71.39
SLR	55.56	80.00	67.78
Group SLR	55.56	**85.00**	70.28
Kernel SLR	**88.89**	50.00	69.45
Kernel group SLR	**88.89**	55.00	**71.95**
SVM	77.78	60.00	68.89
Kernel SVM	77.78	65.00	71.39

occurs the most. For SVM, multiple tasks are incorporated by concatenating all 7 training dictionaries to form a large dictionary $\boldsymbol{X} \in \mathbb{R}^{7n \times p}$. SVM algorithm is then performed on $\boldsymbol{X}$ and voting scheme is employed to assign class label. For the kernel versions, we use RBF kernel with bandwidth selected via cross validation. As one can observe from Table III, our approach outperforms all conventional classifiers.

To further show the efficiency of our approach, we repeat the same experiments using DEC09 data for training and DEC10 data for testing. The classification performances are provided in Table IV and V. In these experiments, we exclude 3th and 4th sensors due to the sensor malfunction in December 09. It can be seen from Table IV that incorporating all 7 sensors yields the best classification accuracy. Table V compares our approach to traditional classifiers. One can observe that MTMV-SRC is comparable to kernel SVM and kernel group SLR. However, we will experimentally show in the next section that by taking clutter noise into account, our MTMV-SRC + N algorithm considerably outperforms all other methods.

C. Deal with arbitrarily large error

This section shows the significance of imposing additional l_1-regularization term into MTMV-SRC model. This regular-

ization is expected to compensate for the unwanted clutter noises in the signals during the data collection. We conduct similar two-class classification experiments with training samples taken from DEC10 data and report results in Table VI. In addition, Table VII shows results with training samples taken from DEC09 data. It can be seen from the tables that by encouraging sparsity for the noise term, we can considerably improve classification accuracy for both classes as well as overall performance.

V. CONCLUSION AND DISCUSSION

In this paper, we propose a novel multi-task multivariate joint structured sparsity-based classification method (MTMV-SRC) for personnel footstep recognition, where the data is collected from nine sensors including acoustic, seismic, PIR and ultrasonic sensors. Our proposed approach shows how to efficiently exploit correlations between sensors measuring the same physical events. Simulation results demonstrate that our method yields highly accurate classification performance and outperforms many classical classifiers such as (group) sparse logistic regression (SLR), support vector machine (SVM) and their kernel versions. Furthermore, we extend our model to deal with large clutter noise, which is indispensable in many practical scenarios. We experimentally illustrate the importance of enforcing another sparse noise regularization into

TABLE VI

CLASSIFICATION ACCURACY (%) FOR TWO CLASS PROBLEM WITH TRAINING SAMPLES TAKEN FROM DEC10, SHOWING THE SUPERIORITY OF TAKING SPARSE NOISE INTO ACCOUNT FOR BOTH MV-SRC AND MTMV-SRC METHODS.

Methods	H	HA	OA
MV-SRC+N sensor 1	81.25	53.33	67.29
MV-SRC+N sensor 2	75.00	60.00	67.50
MV-SRC+N sensor 3	68.75	46.67	57.71
MV-SRC+N sensor 4	**93.75**	6.67	50.21
MV-SRC+N sensor 5	68.75	40.00	53.38
MV-SRC+N sensor 6	56.25	73.33	64.79
MV-SRC+N sensor 7	68.75	66.67	67.71
MV-SRC+N sensor 8	37.75	73.33	55.54
MV-SRC+N sensor 9	68.75	53.33	61.04
MTMV-SRC+N sensors 1-2	81.25	66.67	73.96
MTMV-SRC+N sensors 1-4	56.25	**86.67**	71.46
MTMV-SRC+N sensors 5-7	56.25	**86.67**	71.46
MTMV-SRC+N sensors 1,2,5-9	81.25	73.33	**77.29**

TABLE VII

CLASSIFICATION ACCURACY (%) FOR TWO CLASS PROBLEM WITH TRAINING SAMPLES TAKEN FROM DEC09, SHOWING THE SUPERIORITY OF TAKING SPARSE NOISE INTO ACCOUNT FOR BOTH MV-SRC AND MTMV-SRC METHODS.

Methods	H	HA	OA
MV-SRC+N sensor 1	83.33	35.00	59.17
MV-SRC+N sensor 2	**88.89**	40.00	64.45
MV-SRC+N sensor 5	50.00	**85.00**	67.50
MV-SRC+N sensor 6	61.11	65.00	63.06
MV-SRC+N sensor 7	61.11	45.00	53.06
MV-SRC+N sensor 8	83.33	25.00	54.17
MV-SRC+N sensor 9	61.11	40.00	50.56
MTMV-SRC+N sensors 1-2	83.33	55.00	69.17
MTMV-SRC+N sensors 5-7	61.11	80.00	70.56
MTMV-SRC+N sensors 1,2,5-9	72.22	80.00	**76.11**

MTMV-SRC to remove arbitrarily large noise. This model significantly improves the overall classification accuracy. Lastly, we propose a first-order fast algorithm based on classical alternative direction method to solve aforementioned models.

REFERENCES

[1] D. L. Hall and J. Llinas, "An introduction to multisensor data fusion," *Proceedings of the IEEE*, vol. 85, no. 1, pp. 6–23, 1997.
[2] M. E. Liggins, J. Llinas, and D. L. Hall, *Handbook of Multisensor Data Fusion: Theory and Practice*, 2nd ed. CRC Press, 2008.
[3] P. K. Varshney, "Multisensor data fusion," *Electronics and Communication Engineering Journal*, vol. 9, no. 6, pp. 245–253, 1997.
[4] M. Duarte and Y.-H. Hu, "Vehicle classification in distributed sensor networks," *Journal of Parallel and Distributed Computing*, vol. 64, no. 7, pp. 826–838, 2004.
[5] A. Klausner, A. Tengg, and B. Rinner, "Vehicle classification on multi-sensor smart cameras using feature- and decision-fusion," *IEEE conference on Distributed Smart Cameras*, pp. 67–74, 2007.
[6] E. J. Candès, J.Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *IEEE Transaction on Information Theory*, vol. 52, pp. 5406–5425, 2006.
[7] D. L. Donoho, "Compressed sensing," *IEEE Transaction on Information Theory*, vol. 52, pp. 1289–1306, 2006.
[8] G. Obozinski, B. Taskar, and M. I. Jordan, "Joint covariate selection and joint subspace selection for multiple classification problems," *Journal of Statistics and Computing*, vol. 20, no. 2, pp. 231–252, 2010.
[9] J. Wright, A. Yang, A. Ganesh, S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 31, no. 2, pp. 210–227, 2009.
[10] J. Wright, Y. Ma, J. Mairal, G. Sapiro, T. Huang, and S. Yan, "Sparse representation for computer vision and pattern recognition," *the Proceedings of the IEEE*, vol. 98, no. 6, pp. 1031 – 1044, 2010.
[11] Y. Chen, N. M. Nasrabadi, and T. D. Tran, "Sparse representation for target detection in hyperspectral imagery," *IEEE Journal of Selected Topics in Signal Processing*, vol. 5, no. 3, pp. 629–640, 2011.
[12] H. Zhang, N. M. Nasrabadi, T. S. Huang, and Y. Zhang, "Transient acoustic signal classification using joint sparse representation," *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2011.
[13] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman, "Discriminative learned dictionaries for local image analysis," *Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1 – 8, 2008.
[14] X.-T. Yuan and S. Yan, "Visual classification with multi-task joint sparse representation," *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3493–3500, 2010.
[15] H. O. Park, A. A. Dibazar, and T. W. Berger, "Cadence analysis of temporal gait patterns for seismic discrimination between human and quadruped footsteps," *IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 1749–1752, 2009.
[16] K. M. Houston and D. P. McGaffigan, "Spectrum analysis techniques for personnel detection using seismic sensors," *Proceedings of the SPIE*, vol. 5090, pp. 162–173, 2003.
[17] J. M. Sabatier and A. E. Ekimov, "Range limitation for seismic footstep detection," *Proceedings of SPIE*, vol. 6963, 2008.
[18] M. Yuan and Y. Lin, "Model selection and estimation in regression with grouped variables," *Journal of the Royal Statistical Society: Series B*, vol. 68, no. 1, pp. 49–67, 2006.
[19] J. A. Tropp, "Greed is good: Algorithmic results for sparse approximation," *IEEE Transactions on Information Theory*, vol. 50, no. 10, pp. 2231–2242, 2004.
[20] W. Dai and O. Milenkovic, "Subspace pursuit for compressive sensing signal reconstruction," *IEEE Transactions on Information Theory*, vol. 55, no. 5, pp. 2230–2249, 2009.
[21] T. Blumensath and M. E. Davies, "Iterative hard thresholding for compressed sensing," *Applied and Computational Harmonic Analysis*, vol. 27, no. 3, pp. 265–274, 2009.
[22] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society. Series B*, vol. 58, no. 1, pp. 267–288, 1996.
[23] J. A. Tropp, "Algorithms for simultaneous sparse approximation. part ii: Convex relaxation," *Signal Processing*, vol. 86, no. 3, pp. 589–602, 2006.
[24] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" submitted for publication, 2009.
[25] J. Yang and Y. Zhang, "Alternating direction algorithms for l1 problems in compressive sensing," 2010, preprint.
[26] M. V. Afonso, J. Bioucas-Dias, and M. Figueiredo, "An augmented lagrangian approach to the constrained optimization formulation of imaging inverse problems," 2010, submitted.
[27] S. Engelberg and E. Tadmor, "Recovery of edges from spectral data with noise - a new perspective," *SIAM Journal on Numerical Analysis*, vol. 46, no. 5, pp. 2620 – 2635, 2008.
[28] D. G. Childers, D. P. Skinner, and R. C. Kemerait, "The cepstrum: A guide to processing," *Proceedings of the IEEE*, vol. 65, no. 10, pp. 1428 – 1443, 1977.
[29] L. Meier, S. V. D. Geer, and P. Bhlmann, "The group lasso for logistic regression," *Journal of the Royal Statistical Society: Series B*, vol. 70, no. 1, pp. 53–71, 2008.
[30] L. Cao, J. Luo, F. Liang, and T. S. Huang, "Heterogeneous feature machines for visual recognition," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1095–1102, 2009.
[31] J. Shawe-Taylor and N. Cristianini, *Kernel Methods for Pattern Analysis*. Cambridge University Press, 2004.